

Critical clusters in interdependent economic sectors

A data-driven spectral clustering analysis

Gabriele Oliva^{1,a}, Roberto Setola¹, and Stefano Panzieri²

¹ Unit of Automatic Control, University Campus Bio-Medico di Roma, Via A. del Portillo 21, 00128 Rome, Italy

² University Rome Tre, Department of Engineering, Via della Vasca Navale 79, 00146, Rome, Italy

Received 8 December 2015 / Received in final form 12 February 2016
Published online 26 October 2016

Abstract. In this paper we develop a data-driven hierarchical clustering methodology to group the economic sectors of a country in order to highlight strongly coupled groups that are weakly coupled with other groups. Specifically, we consider an *input-output* representation of the coupling among the sectors and we interpret the relation among sectors as a directed graph; then we recursively apply the *spectral clustering* methodology over the graph, without a priori information on the number of groups that have to be obtained. In order to do this, we resort to the *eigengap* criterion, where a suitable number of groups is selected automatically based on the intensity and structure of the coupling among the sectors. We validate the proposed methodology considering a case study for Italy, inspecting how the coupling among clusters and sectors changes from the year 1995 to 2011, showing that in the years the Italian structure underwent deep changes, becoming more and more interdependent, i.e., a large part of the economy has become tightly coupled.

1 Introduction

In the literature a relevant effort has been spent in finding the most critical elements in a scenario composed of several tightly interconnected economic sectors or *critical infrastructures* (see, among others, [1–3]). Traditional approaches focus on finding the single sectors or infrastructures which are comparatively more vulnerable or critical to the whole system; however, to date, no satisfactory solution has been provided to find *critical groups* of elements or subsystems in the context of *economic input-output analysis* or *critical infrastructure protection*. Indeed, the identification of highly clustered sets of sectors/infrastructures may help understanding the complex relations that exist among the elements that compose such interdependent scenarios. Moreover, finding highly clustered groups from either a structural or functional point of view

^a e-mail: g.oliva@unicampus.it

allows to identify the connections among such groups, which can be regarded as the “weak element of the chain”; such connections are often neglected or shaded by the high degree of coupling of some of the elements.

The analysis of strongly coupled clusters can be used to complement key sector analyses [1,3], allowing to identify *key clusters*.

In this paper, based on the preliminary results in [7], we present a data-driven hierarchical clustering approach to identify groups of tightly interdependent critical infrastructures or economic sectors, taking into account the intensity of the coupling among them. Specifically, we consider an *input-output* representation, where the relations existing in a set of interdependent sectors (infrastructures) is characterized in terms of the economic amount of commodities/services produced by one sector, which is required for the production of commodities/services by another sector (in the case of infrastructures, instead, the relation is expressed in terms of how much the severity of a failure affecting one infrastructure is transferred to the others). These relations are summarized in the *technology matrix* \mathcal{A} [4] (or in the *interdependency matrix* \mathcal{A}^* in the case of infrastructures, which is obtained from \mathcal{A} , by normalization [5]); such a matrix is provided yearly by several institutions, such as BEA (US), Eurostat (EU) or WIOD [6] (<http://www.wiod.org/>).

The above matrix is, in general, full and not symmetric, and can be interpreted as the weighted adjacency matrix of an almost complete directed graph. In this paper, therefore, we seek clusters of strongly coupled sectors by performing a hierarchical spectral clustering decomposition of the graph that corresponds to the technology matrix. More in detail, we rely on a powerful heuristic, namely *eigengap criterion* for the automatic choice of the number of clusters the graph has to be split in, and we iterate the clusterization until all clusters contain just one node.

The procedure yields a hierarchical structure, which can be represented by a tree, or *dendrogram*. In this view, the leaves of the tree are the sectors/infrastructures, while the other nodes represent clusters of infrastructures. We validate the proposed methodology by considering a case study related to the economic input-output data provided by WIOD for Italy in the years from 1995 to 2011.

Specifically, the analysis showed that, in the considered period, the Italian economic sectors have developed a strong clusterization, with the formation of a giant cluster which includes 25 sectors and an economic value of about 10^6 million dollars (while in 1995 the largest cluster counted only 15 sectors, for an economic value of about 4.6×10^5 million dollars). Moreover, the sectors have increased the coupling with other sectors in the same clusters, while the inter-cluster coupling has reduced along the years.

The outline of the paper is as follows: after some preliminary definitions, that conclude this introduction, we review the input-output economic model in Sect. 2; then we review in Sect. 3 the spectral clustering methodology, and we present in Sect. 4 the proposed data-driven hierarchical clustering methodology. Section 5 is devoted to discuss the case study, while some conclusive remarks and future work directions are collected in Sect. 6.

1.1 Preliminaries

Let $\text{diag}(c_1, \dots, c_n)$ be an $n \times n$ diagonal matrix whose diagonal entries are c_1, \dots, c_n . We denote by $|X|$ the number of elements in a set X , and by $q_i(M)$ the right eigenvector associated to the i th smallest eigenvalue $\lambda_i(M)$ of a matrix M . With a slight abuse, we refer to $q_i(M)$ as the i th eigenvector of M .

Let $G = \{V, E, W\}$ be a *graph* with n nodes $V = \{v_1, v_2, \dots, v_n\}$ and e edges $E \subseteq V \times V$, where $(v_i, v_j) \in E$ captures the existence of a link from node v_i to node

v_j . The $n \times n$ matrix W is the *weighted adjacency matrix*, whose elements $w_{ij} \neq 0$ iff $(v_j, v_i) \in E$; w_{ij} is the *weight* of the edge (v_j, v_i) . A weighted graph is said to be *undirected* if $(v_i, v_j) \in E$ whenever $(v_j, v_i) \in E$ and $w_{ij} = w_{ji}$, and is said to be *directed* otherwise.

The *in-degree* d_i^{in} of a node v_i is the sum of the weight of its incoming edges, i.e., $d_i^{in} = \sum_{j=1}^n w_{ij}$, while the *out-degree* d_i^{out} is the number of its outgoing edges, i.e., $d_i^{out} = \sum_{j=1}^n w_{ji}$. For undirected graphs, it always holds that $d_i^{in} = d_i^{out}$, and in this case it is simply referred to as the *degree* d_i of node v_i .

A *path* P_{ij} over a graph $G = \{V, E, W\}$, starting from a node $v_i \in V$ and ending in a node $v_j \in V$, is a subset of links in E that connects v_i and v_j without creating loops.

A graph is *connected* if for each pair of nodes v_i, v_j there is a path over G that connects them without necessarily respecting the edge orientation, while it is *strongly connected* if the path respects the orientation of the edges. It follows that every undirected connected graph is also strongly connected.

A *tree* T is a connected acyclic undirected graph; it is possible to specify a node v_i as the *root* of the tree. A leaf in a tree T rooted at a node v_i is a node $v_j \neq v_i$ whose degree is $d_j = 1$. The *parent* v_j of a node v_i in a tree is the neighbor of v_i lies in the path from v_i to the root node (the root node does not have a father), while a node v_j is a *son* of v_i if v_i is the father of v_j (a node in a tree can have, in general, several sons). The depth of a node v_i in a tree T is the length of the path connecting the root node and v_i , in terms of number of links in the path (the root node has zero depth).

The *Laplacian matrix* L of a graph G is given by

$$L = D - W,$$

where

$$D = \text{diag}(d_1^{in}, \dots, d_n^{in}),$$

while the *normalized Laplacian matrix* is given by

$$L_{norm} = D^{-1/2} L D^{-1/2}.$$

The eigenvalues of the (normalized) Laplacian matrix satisfy

$$0 = \lambda_1(L) \leq \lambda_2(L) \leq \dots \leq \lambda_n(L),$$

and, in the undirected graph case, they are all real. Moreover, for undirected graphs, the multiplicity of the eigenvalue 0 coincides with the number of *connected components* of G , hence the multiplicity is 1 if the graph G is connected.

2 Input-output modeling of coupled economic sectors

The Input-Output model [4] is a linear model that represents how much each sector in an economy has to produce in order to meet an external demand, highlighting the relations existing among the economic sectors. In the input-output approach the product of each sector is expressed in monetary value (e.g., million dollars), and the

model captures the relations between the sectors in terms of the amount of product of a given sector i required by sector j to produce one unit of product (e.g., one million dollars worth). Let

$$X = [X_1, \dots, X_n]^T$$

be the vector containing the total economic output of the different sectors for a given year and let

$$\Delta = [\Delta_1, \dots, \Delta_n]^T$$

be the amount of external demand for each sector (each X_i and Δ_i are expressed in million dollars); moreover, let Z be the $n \times n$ *input-output* matrix, whose entries Z_{ij} represent the amount of product of sector i (in million dollars) that is required by sector j to produce its product, in a given year. Matrix Z is typically provided yearly by several institutions, such as BEA (US) or Eurostat (EU).

Let the *technology matrix* be an $n \times n$ matrix \mathcal{A} , whose coefficients \mathcal{A}_{ij} represent the fraction of production of sector i that is required to produce one unit of the product of sector j .

The technology matrix is obtained from the input-output matrix, normalizing each entry Z_{ij} by X_i , i.e.,

$$\mathcal{A} = \text{diag}(X_1, \dots, X_n)^{-1} Z,$$

and the input-output model is given by

$$X = \mathcal{A}X + \Delta.$$

2.1 Inoperability input-output model

In this subsection we briefly review an extension of the above model which, although being out of the scope of present paper, is given for completeness and in order to give an idea of possible future work directions.

In [5], the above model is extended to represent the interdependency relations existing among coupled *critical infrastructures*; in this view, the *inoperability* Q_i of an infrastructure i is introduced as its percentage of malfunctioning, while the exogenous disturbance Δ_i^* can be regarded as the severity of an outage (natural or man-made) affecting the i th infrastructure. Specifically, the model initially considers how an imbalance $\hat{\Delta}$ of external demand affects the variation of production \hat{X} , i.e.

$$\hat{X} = \mathcal{A}\hat{X} + \hat{\Delta},$$

and then the inoperability Q is obtained from \hat{X} by normalization, i.e.,

$$Q = \text{diag}(X_1, \dots, X_n)^{-1} \hat{X}$$

resulting in

$$Q = \mathcal{A}^*Q + \Delta^*,$$

where

$$\mathcal{A}^* = \text{diag}(X_1, \dots, X_n)^{-1} \mathcal{A} \text{diag}(X_1, \dots, X_n)$$

and

$$\Delta^* = \text{diag}(X_1, \dots, X_n)^{-1} \hat{\Delta}.$$

2.2 Input-output model as a graph

As discussed in the introduction, we are interested in grouping the economic sectors of a nation in order to create clusters characterized by a strong interrelation among elements belonging to the cluster and by a limited interaction with elements outside the cluster. We are, moreover, interested in decomposing further each cluster, in order to gain insights on the structure of the coupling among the sectors, clusters and sub-clusters.

If we interpret the coefficients \mathcal{A}_{ij} as the weights of the links in an almost complete graph where the nodes coincide with the sectors, our problem becomes how to group the nodes in the graph such that the sum of the weights of the links inside the group is comparatively high, while the sum of the weights of the links that connect different groups is comparatively low.

We show in the next section how to accomplish such a task via spectral clustering methodologies, while we present a hierarchical clustering approach based on spectral clustering in Sect. 4.

Notice that, in order to focus on the interaction among sectors, in the following we do not consider the terms \mathcal{A}_{ii} , which would correspond to self-links (i.e., from a node v_i to itself).

3 Spectral clustering

This section is devoted to illustrate the spectral clustering methodology adopted in this paper, while next section aims at presenting the proposed hierarchical clustering approach.

In the context of *spectral clustering*, we want to partition the nodes in a weighted connected and undirected graph $G = \{V, E, W\}$ into k groups such that the weights of the links inside a group are large, while the weights of the links that cross the boundary of the group are small. Moreover, we want the partitions to be as balanced as possible, in terms of number of nodes assigned to each partition.

3.1 Two clusters

For $k = 2$, the problem is known as the *Normalized Minimum-Cut* problem, and we want to divide the nodes in two groups A and \bar{A} , minimizing the objective function

$$Ncut(A, \bar{A}) = cut(A, \bar{A}) \left(\frac{1}{vol(A)} + \frac{1}{vol(\bar{A})} \right)$$

where

$$cut(A, \bar{A}) = \sum_{i \in A, j \in \bar{A}} w_{ij}$$

is the *cut* between A and \bar{A} , while

$$vol(A) = \sum_{i \in A} d_i$$

is the *volume* of the partition A .

Finding $cut(A, \bar{A})$ is an easy task, and in the literature there are efficient algorithms [8, 9]. The normalized cut, conversely, is much harder to solve exactly. In [10],

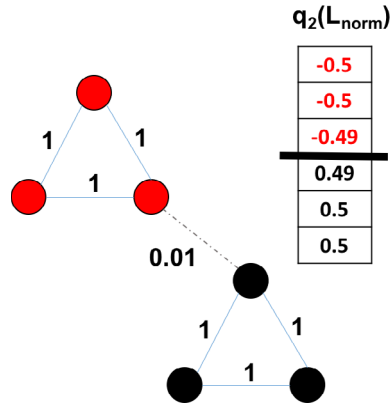


Fig. 1. Example of spectral clustering for $k = 2$ groups over an undirected weighted graph with $n = 6$ nodes. The eigenvector $q_2(L_{norm})$ provides a clear division in two sets A (in red, negative entries) and \bar{A} (in black, positive entries).

however, a very good approximated solution is given, based on the normalized laplacian matrix L_{norm} .

Specifically, the algorithm in [10] calculates the second eigenvector $q_2(L_{norm})$, and assigns each node v_i to the clusters A or \bar{A} based on the sign of the corresponding component of $q_2(L_{norm})$; an example of the above procedure is given in Fig. 1.

3.2 More than two clusters

In [11] the above approach is extended to $k > 2$; in this case we want to find a normalized cut for k disjoint partitions A_1, \dots, A_k , i.e., we want to minimize

$$Ncut(A_1, \dots, A_k) = \sum_{i=1}^k \frac{cut(A_i, \bar{A}_i)}{vol(A_i)},$$

where $\bar{A}_i = V - A_i$ for all $i = 1, \dots, k$.

Similarly to the case for $k = 2$, in [11] the matrix

$$U = [q_2(L_{norm}), \dots, q_k(L_{norm})]$$

is considered, and the i -th row of U is associated to the i -th node in the graph. The above association is, therefore, a projection of the nodes of the graph G in \mathbb{R}^{k-1} . Notice that it is not immediate to partition the projected points based on U , and the projected points must be clustered in k groups using data clustering techniques such as the k -means algorithm [12] (an example of the above procedure for $k = 3$ is given in Fig. 2).

3.3 Extension to directed graphs

If the graph G is directed, the above techniques may fail [13]. In the literature, several methods [13–15] have been proposed to cast the laplacian matrix of a directed graph into a symmetric laplacian matrix that takes into account the original directed links.

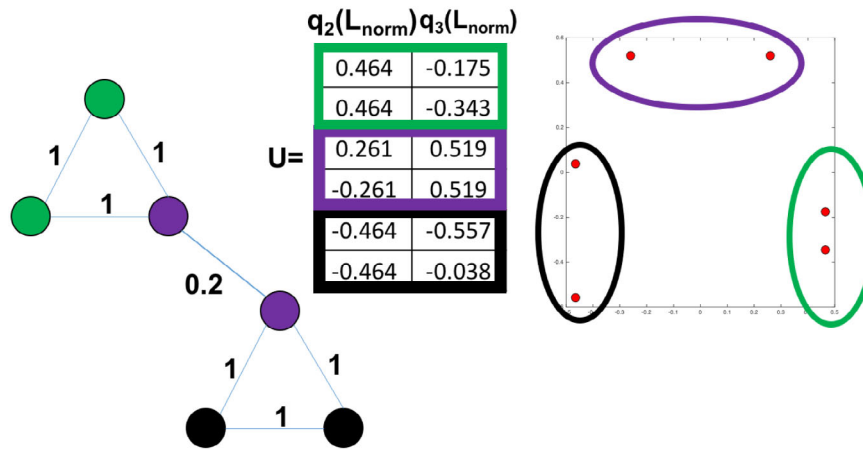


Fig. 2. Example of spectral clustering for $k = 3$ groups over an undirected weighted graph with $n = 6$ nodes. The eigenvectors $q_2(L_{\text{norm}})$ and $q_3(L_{\text{norm}})$ are used to map the nodes of the graph in \mathbb{R}^2 , and then are clustered via the k-means algorithm. The clusters are shown in green, purple and black.

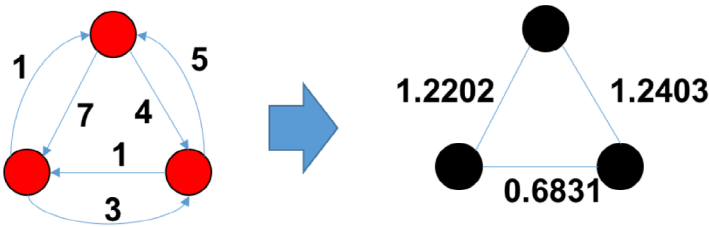


Fig. 3. Example of conversion of the weights of a directed graph (left) into weights of an undirected graph, following the approach in [16].

Let us discuss the approach in [16], which results in simple computations. Such a method implicitly assumes that each node has both nonzero in-degree and nonzero out-degree.

Let us consider the *in-degree* and *out-degree* matrices D_i and D_o , defined as

$$D_i = \text{diag}(d_1^{\text{in}}, \dots, d_n^{\text{in}}), \quad D_o = \text{diag}(d_1^{\text{out}}, \dots, d_n^{\text{out}}).$$

In order to take into account both the in-degree and the out-degree of the nodes, in [16] the matrix $\Phi_{io} = (D_{in}D_{out})^{1/2}$ is introduced, and an undirected weight matrix W_{io} is obtained as

$$W_{io} = \Phi_{io}^{-1/2}(W + W^T)\Phi_{io}^{-1/2}.$$

Then, the spectral clustering is applied to the laplacian matrix L_{io} obtained from W_{io} instead of W , i.e.,

$$L_{io} = D_{io} - W_{io},$$

where D_{io} is the diagonal matrix whose entries are equal to the sum of the rows of W_{io} ; an example is given in Fig. 3. Notice that, as a result of the above procedure, W_{io} is symmetric, and therefore it represents the weighted adjacency matrix of an undirected graph.

Let us conclude the section by discussing a way to choose automatically the value of k , which is a fundamental point for the developments of this paper.

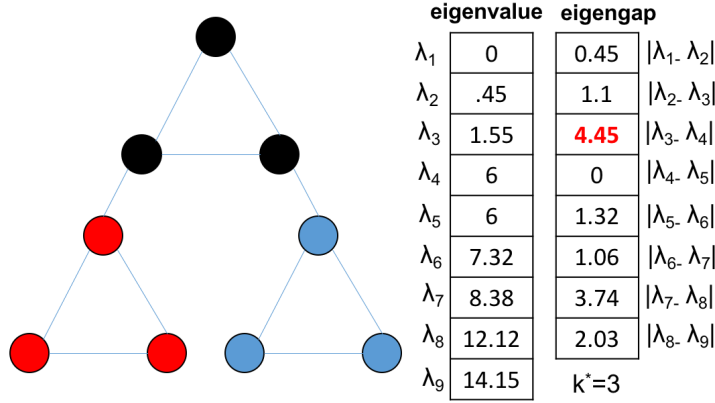


Fig. 4. Example of automatic choice of k via the eigengap heuristic over a graph G with $n = 9$ nodes and unitary weights. The number $k = 3$ clusters is chosen as the maximum argument of the eigengap. The three clusters thus obtained are shown in black, red and blue.

3.4 Automatic choice of k

The true weak point of any clustering technique is that the number k of clusters must be specified by the user, which must have a priori information on the structure of the graph in order to select a suitable number of clusters.

In the case of spectral clustering, however, we can use a simple, yet powerful heuristic approach to derive k automatically [17,18] (we report an example in Fig. 4). Specifically, we choose the value k^* that maximizes the *eigengap* of the Laplacian matrix L_{norm} , i.e.,

$$k^* = \arg \max_{k=1, \dots, n-1} \{|\lambda_k(L_{norm}) - \lambda_{k+1}(L_{norm})|\}.$$

An intuitive explanation for this choice comes from the fact that, as discussed in Sect. 1.1, in the ideal case of k^* completely disconnected clusters the zero eigenvalue of L_{norm} has multiplicity k^* , and there is a relevant gap between the k^* th eigenvalue of L_{norm} (which is zero) and the $(k^* + 1)$ th one. Analogously, when the graph is composed of k^* dense clusters and the clusters are linked via links with small total weights, the eigengap is likely to reach its maximum at k^* .

In the next section we present a hierarchical clustering approach based on the above heuristic criterion.

4 Data-driven hierarchical clustering

In this section we develop a data-driven hierarchical clustering methodology that does not rely on a priori knowledge about the number of groups; instead, it is based on the eigengap criterion discussed in the previous section.

As discussed above, the eigengap heuristic is an effective way to partition the nodes of a graph G in a number k^* groups which is not known a priori, but depends on the topology and on the intensity of the coupling among the nodes.

If we recursively execute the procedure over the clusters, until all clusters contain just one node, we obtain a hierarchical clustering. Such a clustering can be represented by a *dendrogram* (i.e., a tree), where the leaves are the nodes of the original graph G ,

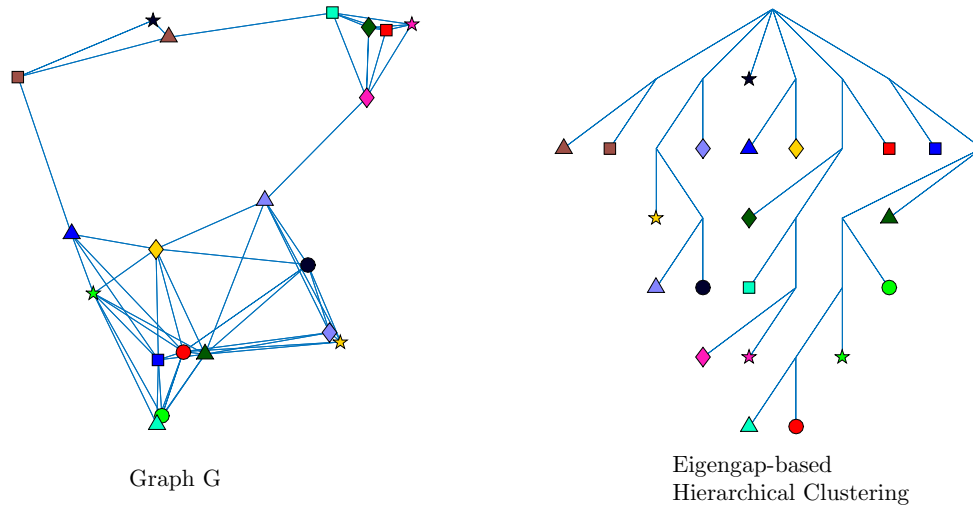


Fig. 5. Example of hierarchical clustering based on the eigengap heuristic. The left plot shows the graph G ($n = 20$ nodes, unitary weights) while the right plot shows the tree representing the hierarchical clustering (the tree has $m = 34$ nodes, and the leaves coincide with the n nodes in G).

while the other nodes in the tree represent the different clusters and sub-clusters (in this view, the root node can be regarded as the set containing all the nodes of G).

Notice that, while traditional hierarchical clustering approaches [19,20] recursively decompose the groups in two groups, resulting in a binary tree, here the number of cluster is not specified a priori, but it depends on the structure of the graph/clusters (i.e., we seek strongly coupled communities that are loosely coupled with the other communities).

Figure 5 shows an example of the above procedure. In the figure, we consider an undirected connected graph G with unitary weights and we cluster the nodes in the graph by means of the eigengap approach. Specifically, the eigengap heuristic yields $k^* = 6$ groups, of which just one is a singleton (black star). The other clusters are, therefore, decomposed further via the same approach as above, and so on until all clusters are composed of just one node. It can be noted that, while the node corresponding to a black star is immediately isolated from the other infrastructures, the nodes represented by the green triangles and the red circle belong to a big cluster (6 nodes) and remain in the cluster after several rounds of partitioning (it takes 6 rounds to obtain a singleton), meaning that these nodes are at the “core” of the partition and are quite influent on the other nodes in the partition. Also, the partition they belong to at the first round is quite coupled, as it loses just one element at each further round of division (i.e., it is decomposed in a singleton and a set containing all nodes but the one in the singleton). Let us now discuss some coupling indicators that stem from the above intuitions.

4.1 Coupling indicators

Let us consider the following indicators related to the structure of the clusters obtained at the first round of division (i.e., the sons of the root node in the dendrogram),

as such partitions represent the first and more evident clusterization of the graph G . Specifically, we take into account:

- the cardinality $|A_i|$ of each cluster A_i (i.e., the number of nodes of G that belong to the cluster A_i);
- the total sum

$$\zeta_i = \sum_{v_a, v_b \in A_i} w_{ab}$$

of the weights within each cluster A_i ;

- the sum

$$\eta_{ij} = \sum_{v_a \in A_i, v_b \in A_j} w_{ab}$$

of the weights of the links that connect each pair of clusters A_i, A_j .

In particular, the cardinality of a cluster A_i and ζ_i provide a measure of the degree of coupling in a given cluster, while η_{ij} is a measure of the coupling among two clusters. Notice that values of η_{ij} that are remarkably smaller than ζ_i and ζ_j suggest that the clustering procedure has been successful.

5 Case study

In this section we consider the input-output matrices Z provided by WIOD [6] for Italy, on a yearly base from the year 1995 to 2011 (the monetary values are reported in current prices as of 2015). For space issues, we do not report the coefficients of the matrices Z ; the interested reader can access the dataset at <http://www.wiod.org/>. We consider $n = 34$ sectors, as reported in Table 1, and we apply the data-driven hierarchical clustering methodology presented in Sect. 4.

Figures 6 and 7 show the dendrograms representing the results of the hierarchical clustering for the years 1995 and 2011, respectively. The leaves in each dendrogram are reported with black triangles, and the identifier of the corresponding sector is shown next to the triangles. The clusters, conversely, are reported via blue circles and the cardinality of the cluster is reported next to the circle in curly brackets. For each edge in the dendrogram, the value ζ_i associated to the lowermost endpoint of the edge is reported (we show the monetary value, in terms of the corresponding entries of the Z matrix, including the diagonal entries).

According to Fig. 6, it can be noted that the sectors are clustered in coherent groups: we have that inland and water transportations are grouped together in cluster obtained at the first round, while another cluster contain sectors 15 and 26, which are both related to transportations. Moreover, sector 17 and 8 (both related to energy) are grouped together at the first round. As for the cluster containing 8 sectors, it can be noted that most of them are related to the public sector. Inspecting further the structure of the cluster composed of 18 sectors, it can be noted that the subclusters are related each to manufacturing, retail, sales or health and chemicals.

As for the dendrogram in the year 2011 (Fig. 7), it should be noted that, although some sectors (for instance sectors 24 and 25) change slightly their depths and the composition of the clusters they belong to, other sectors change significantly, e.g., 17 (Electricity Gas and power) is now part of the bigger cluster and is at depth 4 (in 1995 its depth was 2).

Figure 8 shows a comparison of the clusters obtained at the first level in 1995 and 2011, in terms of cardinality and total weights ζ_i (in monetary value). As evident also by comparing the first level of the dendrogram in Figs. 6 and 7, it can

Table 1. Sectors considered in the case study. Source: WIOD [6] (<http://www.wiod.org/>)

ID	Denomination
1	Agriculture, Hunting, Forestry and Fishing
2	Mining and Quarrying
3	Food, Beverages and Tobacco
4	Textiles and Textile Products
5	Leather, Leather and Footwear
6	Wood and Products of Wood and Cork
7	Pulp, Paper, Paper, Printing and Publishing
8	Coke, Refined Petroleum and Nuclear Fuel
9	Chemicals and Chemical Products
10	Rubber and Plastics
11	Other Non-Metallic Mineral
12	Basic Metals and Fabricated Metal
13	Machinery, Nec
14	Electrical and Optical Equipment
15	Transport Equipment
16	Manufacturing, Nec; Recycling
17	Electricity, Gas and Water Supply
18	Construction
19	Sale, Maintenance and Repair of Motor Vehicles and Motorcycles; Retail Sale of Fuel
20	Wholesale Trade and Commission Trade, Except of Motor Vehicles and Motorcycles
21	Retail Trade, Except of Motor Vehicles and Motorcycles; Repair of Household Goods
22	Hotels and Restaurants
23	Inland Transport
24	Water Transport
25	Air Transport
26	Other Supporting and Auxiliary Transport Activities; Activities of Travel Agencies
27	Post and Telecommunications
28	Financial Intermediation
29	Real Estate Activities
30	Renting of M&Eq and Other Business Activities
31	Public Admin and Defense; Compulsory Social Security
32	Education
33	Health and Social Work
34	Other Community, Social and Personal Services

be noted that while in 1995 the clustering yields 6 groups with smaller cardinalities (from 1 to 15), in 2011 we obtain a much bigger cluster of 25 sectors and 4 more clusters with cardinality between 1 and 4; hence, the degree of coupling among the sectors is significantly increased. As for the weight ζ_i of the clusters, it can be noted that in 2011 the economic value of the biggest cluster nearly tripled with respect to 1995.

Figures 9 and 10 show the value of the total weights η_{ij} between pairs of clusters identified at the first level, for the years 1995 and 2011, respectively (The nodes in the figures represent the clusters and for each cluster A_i the cardinality $|A_i|$ and economic value ζ_i are shown in curly brackets, i.e., $\{|A_i| \mid \zeta_i\}$). Indeed, according to Fig. 10,



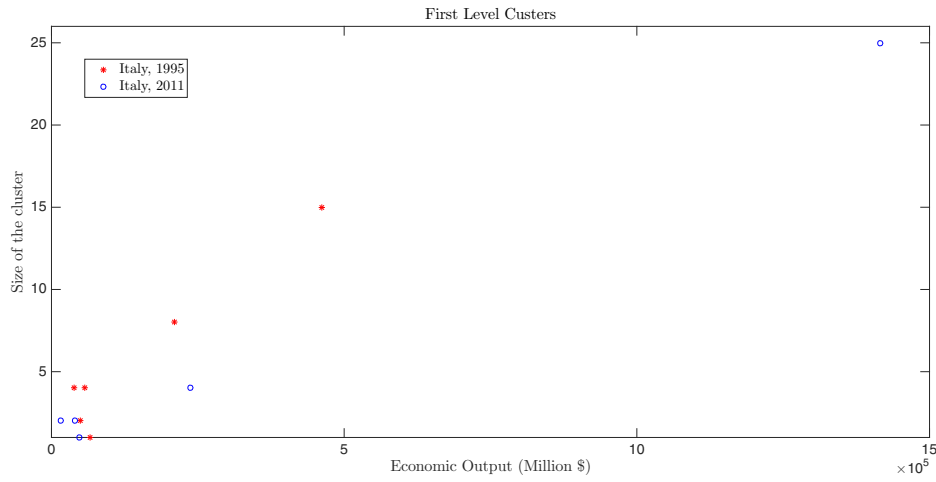


Fig. 8. Comparison of the clusters obtained at the first round of the hierarchical clustering (i.e., the leaves of the root node in the corresponding dendrogram) in terms of cardinality and economic output.

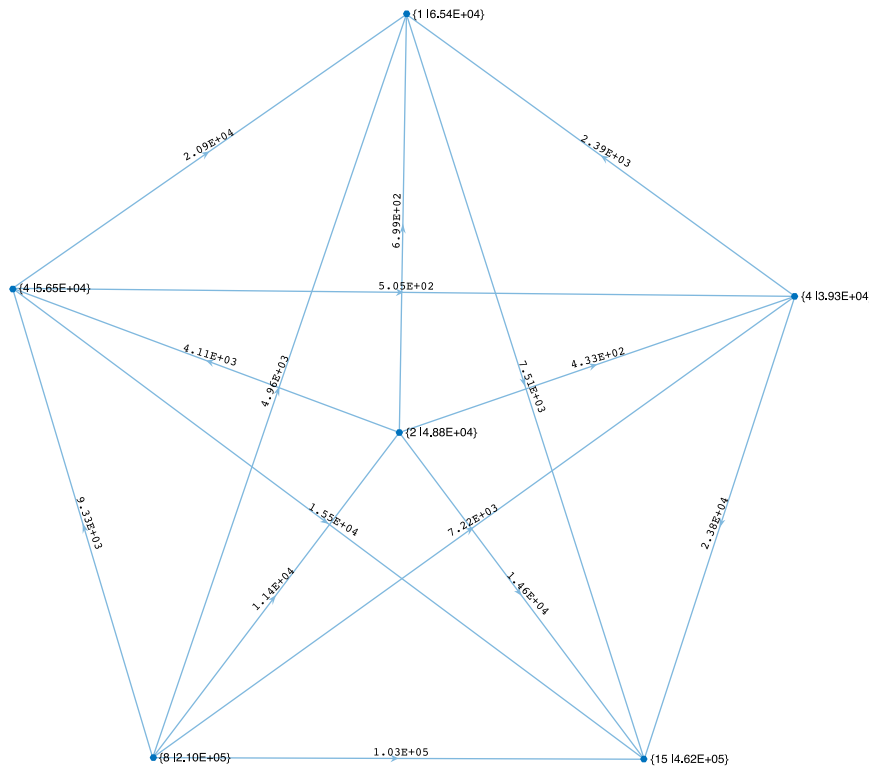


Fig. 9. Graph showing the total weights η_{ij} of the links connecting each pair of clusters obtained at the first round of the hierarchical clustering (i.e., the leaves of the root node in the corresponding dendrogram), for Italy in the year 1995. The cardinality and the sum ζ_i of the weights within each cluster (expressed in million dollars) are reported in curly brackets next to the corresponding node.

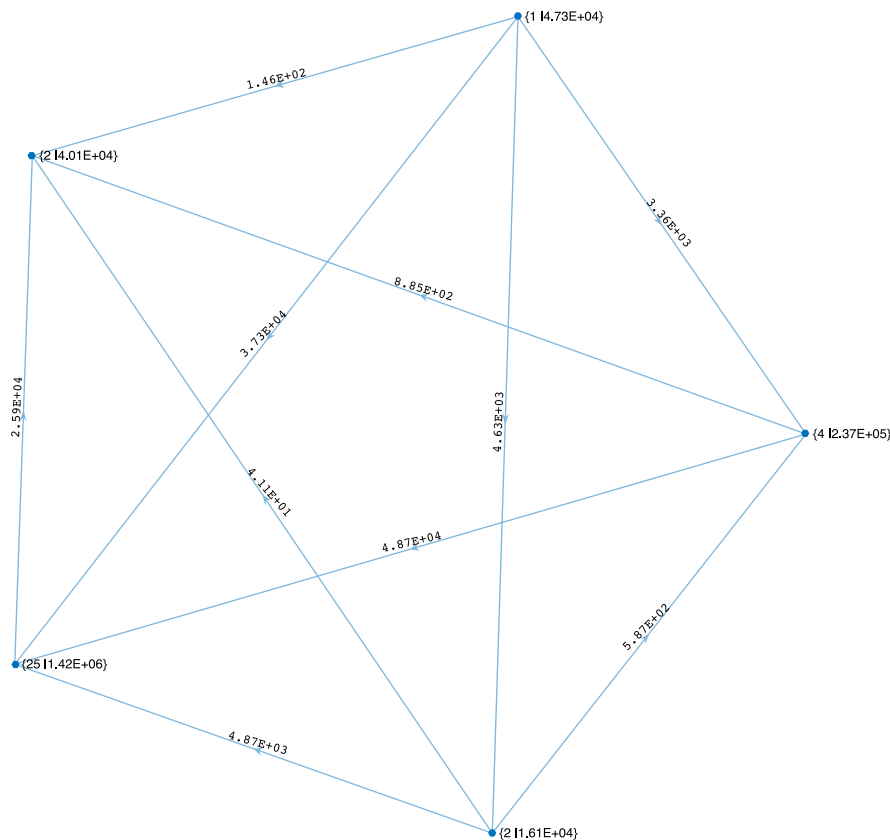


Fig. 10. Graph showing the total weights η_{ij} of the links connecting each pair of clusters obtained at the first round of the hierarchical clustering (i.e., the leaves of the root node in the corresponding dendrogram), for Italy in the year 2011. The cardinality and the sum ζ_i of the weights within each cluster (expressed in million dollars) are reported in curly brackets next to the corresponding node.

it can be noted that (except for few pairs of clusters) the weight η_{ij} is considerably smaller than the cluster weights ζ_i and ζ_j (up to two orders of magnitude smaller); this situation is much more evident in Fig. 10 (i.e., for the year 2011), where η_{ij} is between one and three orders of magnitude smaller than ζ_i and ζ_j . These results, together, suggest that the clustering thus obtained is able to capture the actual clusterization among economic sectors.

Figure 11 reports, plotted against the years, the maximum cluster cardinality and the maximum value of the total economic output (expressed in monetary value) of the elements in a cluster (i.e., the sum of the total outputs of the sectors in the cluster). It can be noted that (except for the year 2004 where there is an evident but momentary reduction) there is a constant increase in both the maximum cluster cardinality and the maximum total output of a cluster.

The results in this section, together, suggest that from one side the Italian economic sectors have increased their mutual coupling and, from another point of view, that there has been a strong clusterization of the sectors, which have increased their coupling within the cluster, while the inter-cluster coupling has indeed reduced of a relevant amount.

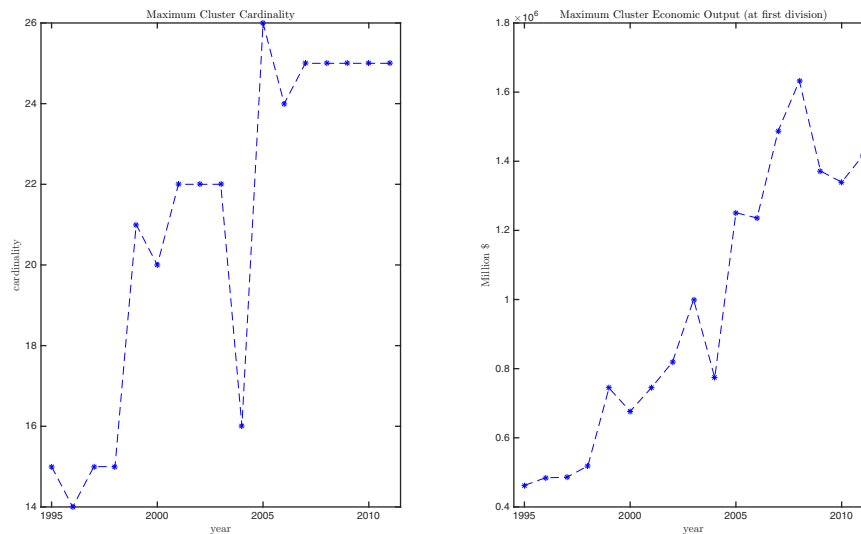


Fig. 11. Global indicators of the overall degree of coupling, plotted against the years: the leftmost plot reports the maximum cardinality of a cluster, while the rightmost plot shows the maximum total economic output of a cluster (i.e., the sum of the total output of the sector composing it, expressed in million dollars).

6 Conclusions and future work directions

In this paper we provide a novel approach to identify clusters of strongly coupled sectors in economies represented via the input-output formalism. Specifically, we resort to a spectral clustering decomposition where the number of groups is not known a priori, and we iterate the process on the clusters until we obtain a hierarchical clustering structure. The proposed methodology is validated with respect to a case study where the economic sectors in Italy are considered from the year 1995 to 2011.

Future work will be aimed to apply the methodology to a broader case study, considering critically different data sources. We will, moreover, inspect the possibility to apply the approach to the case of coupled critical infrastructures, in order to provide a useful support to decision makers that have to decide how to prioritize the protection of such infrastructures. A further envisaged work direction is to frame the results obtained at the national level in the general context of globalization, by comparing the clustering pattern obtained over the years against global data and indicators, such as world input-output tables or export flows (as done in [21]).

The research activity for this work was performed under the CIPRNet Project funded by the European Union's Seventh Framework Programme for Research, Technological Development and Demonstration under Grant no. 312450.

References

1. H.B. Chenery, T. Watanabe, *Econometrica* **26**, 487 (1958)
2. R. Setola, *Int. J. Sys. Sys. Eng.* **2**, 38 (2010)
3. G. Oliva, R. Setola, K. Barker, *Reliability*, *IEEE Transactions* **63**, 42 (2014)
4. W. Leontief, *Input-Output Economies* (Oxford University Press, New York, 1966)
5. Y.Y. Haimes, B.M. Horowitz, J.H. Lambert, J.R. Santos, C. Lian, K.G. Crowther, *J. Infrastructure Sys.* **11**, 67 (2005)

6. M.P. Timmer, E. Dietzenbacher, B. Los, R. Stehrer, G.J. Vries, *Rev. Int. Economics* **23**, 575 (2015)
7. G. Oliva, S. Panzieri, R. Setola, *10th International Conference on Critical Information Infrastructures Security* (Berlin, Germany, October 5–7, 2015 (CRITIS2015)), (to appear)
8. J. Hao, J.B. Orlin, *J. Algorithms* **17**, 424 (1994)
9. M.S. Levine, *Experimental Study of Minimum Cut Algorithms* (1997)
10. J. Shi, J. Malik, *IEEE Transactions on Pattern Anal. Machine Intelligence* **22**, 905 (2000)
11. S.X. Yu, J. Shi, *Proceedings Ninth IEEE International Conference on. Computer Vision*, (IEEE, 2003), p. 313
12. J. MacQueen et al., *Proc. fifth Berkeley Symp. Math. Stat. Probability* **1**, 14 (1967)
13. Y. Kim, S.-W. Son, H. Jeong, in *Complex Sciences* (Springer, 2009), p. 2050
14. E.A. Leicht and M.E. Newman, *Phys. Rev. Lett.* **100**, 118703 (2008)
15. Y. Kim, S.-W. Son, H. Jeong, *Phys. Rev. E*, **81**, 016103 (2010)
16. A. Mirzal, M. Furukawa, *International Conference on Electronics and Information Engineering (ICEIE)*, 2010, Vol. 1. (IEEE, 2010), p. V1
17. F.R. Chung, in *Spectral Graph Theory* (AMS, Providence, 1997), Vol. 92
18. B. Mohar, *Some applications of Laplace eigenvalues of graphs* (Springer, 1997)
19. L. Kaufman, P.J. Rousseeuw, *Finding groups in data: an introduction to cluster analysis* (John Wiley & Sons, 2009), Vol. 344
20. T. Hastie, R. Tibshirani, J. Friedman, *The elements of statistical learning* (Springer, 2009)
21. G. Caldarelli, M. Cristelli, A. Gabrielli, L. Pietronero, A. Scala, A. Tacchella, *PloS one* **7**, e47278 (2012)

Open Access This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.